

Developing a Big Data Ecosystem for Imaging, Spectroscopy and Diffraction at Brookhaven National Laboratory

Eric Stach
Brookhaven National Laboratory

With the advent of ultra-sensitive new detectors and bright new photon sources, scientific User Facilities are generating unprecedented levels of rich data sets. At the Brookhaven National Laboratory, we are home to a new third-generation light source, as well as a nanoscale research center with two high-speed direct electron detectors on electron microscopes. Presently, we are generating several petabytes (Pb) of image, spectroscopy and diffraction data per year, but as the light source beamline build out, we can readily foresee data streams as large as 20 Pb/yr. This talk will describe how the Laboratory is building a Big Data Ecosystem, in an effort to fully exploit the rich information trove that this data represents. Firstly, using an example from the Center for Functional Nanomaterials, we will describe how we deal with a 3Gb/s data stream from our environmental transmission electron microscope by exploiting the capabilities of the RHIC/Atlas Computing Facility, which is primarily focused on managing the data from the high-energy physics side of the laboratory. We will discuss the use of high-throughput computing for data reduction, and the distribution of User data via Globus Online and the ESNet. Thereafter, we will describe how BNL is building capabilities for both near real time data analysis at multiple beamlines and microscopes, in order to ensure that valuable facility time is best utilized. I will also describe how we foresee an integrated data analytics and data mining effort which will allow deep integration of multiple data streams, revolving around experiments conducted with samples 'in a working condition' (i.e. operando). The presentation will highlight the many challenges that the creation of this Ecosystem holds, as well as point towards the ways that the full utilization of large, rich data streams can advance materials research.